Brain network science modelling of sparse neural networks enables Transformers and LLMs to perform as fully connected

Yingtao Zhang[1], Diego Cerretti[1], Jialin Zhao[1], Wenjing Wu[1], Ziheng Liao[1]
Umberto Michieli[2] & Carlo Vittorio Cannistraci[1]

[1]Center for Complex Network Intelligence (CCNI), Tsinghua University; [2]University of Padova
*Correspondence to: Carlo Vittorio Cannistraci <kalokagathos.agon@gmail.com>.

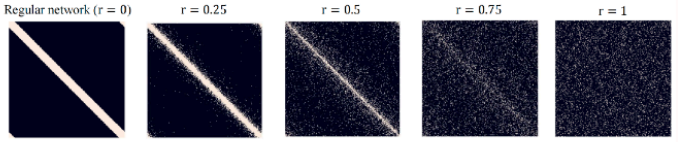
Homepage Paper Code

Published at NeurIPS 2025.

Epitopological local minima. In the context of dynamic sparse training methods, we define an epitopological local minima (ELM) as a state where the sets of removed links and regrown links exhibit a significant overlap.

Cannistraci-Hebb soft removal and regrowth. To address ELM, we adopt a probabilistic approach where the process of both regrowth and removal can be viewed as sampling from a $\{0, 1\}$ multinomial distribution, with the score assigned by either removal metrics or link prediction scores, introducing a "soft sampling" mechanism.

Bipartite Receptive Field network modelling.



Node-based link regrowth. In the original CHT framework, the time complexity of the path-based CH3-L3 metric is $O(N \cdot d^3)$, where N is the number of nodes and d is the network's average degree. To address this issue, we introduce a more efficient, node-based paradigm, **CH2-L3n**, that reduces the time complexity to $O(N^3)$ while maintaining performance.

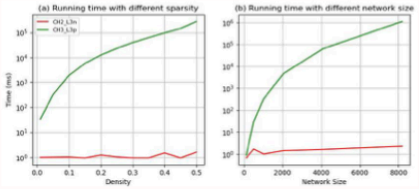


Figure 1: **Runtime Performance Evaluation** of node-based and path-based methods across varying densities and network sizes.

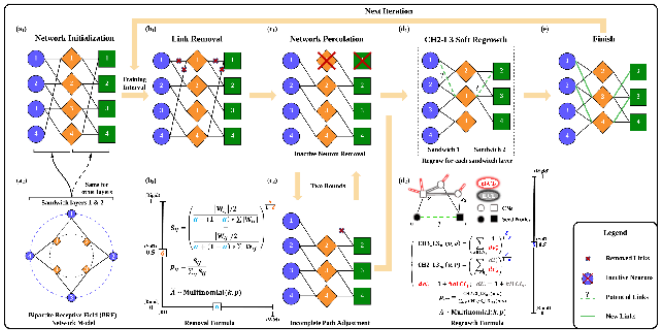


Figure 2: Illustration of the CHTs process. One training iteration follows the path of (a1) Network initialization $->$ (b1) Link removal $->$ (c1) Removal of inactive neurons caused by link removal $->$ (c2) Adjust and remove incomplete links caused by inactive neuron removal $->$ (d1) Link Regrowth $->$ (e)Finished state of the network after one iteration.

(CHTss) Sigmoid gradual decrease density. We propose a sigmoid-based gradual density decrease strategy, defined as:

$$s_t = s_i + (s_i - s_f)\left[\frac{1}{1 + e^{-k\left(t - \frac{t_f - t_0}{2}\right)}}\right], \tag{1}$$

This strategy ensures a smoother initial pruning phase, allowing the model to warm up and stabilize before undergoing significant pruning, thereby enhancing training stability and performance.

Experiments

Table 1: Performance comparison on machine translation tasks of Multi30k, IWSLT, and WMT with varying sparsity levels.

Method	Multi30k		IWSLT		WMT	
	95%	90%	95%	90%	95%	90%
FC	31.38 ± 0.38		24.48 ± 0.30		25.22	
SET	28.99 ± 0.28	29.73 ± 0.10	18.53 ± 0.05	20.13 ± 0.08	20.19 ± 0.12	21.52 ± 0.28
RigL	29.94 ± 0.27	30.26 ± 0.34	20.53 ± 0.21	21.52 ± 0.15	20.71 ± 0.21	22.22 ± 0.10
CHT	27.79	28.38	18.59	19.91	19.03	21.08
CHTs	28.94 ± 0.57	29.81 ± 0.37	21.15 ± 0.10	21.92 ± 0.17	20.94 ± 0.63	22.40 ± 0.06
MEST	28.89± 0.26	30.04 ± 0.52	19.56 ± 0.10	21.05 ± 0.21	20.70 ± 0.07	22.22 ± 0.10
GMP	30.51 ± 0.82	30.49 ± 0.40	22.76 ± 0.82	22.82 ± 0.53	22.47 ± 0.10	23.37 ± 0.08
GraNet	31.31 ± 0.31	31.62 ± 0.48*	22.53 ± 0.12	22.43 ± 0.09	22.51 ± 0.21	23.46 ± 0.09
CHTss	**32.03 ± 0.29***	**32.86 ± 0.16***	**24.51 ± 0.02***	**24.31 ± 0.04**	**23.73 ± 0.43**	**24.61 ± 0.14**

Table 2: **Validation perplexity (↓)** on OpenWebText using LLaMA-60M, 130M, and 1B across varying sparsity levels.

Method	LLaMA-60M				LLaMA-130M				LLaMA-1B
	70%	80%	90%	95%	70%	80%	90%	95%	70%
FC	26.56				19.27				14.62
SET	31.77	30.69	35.26	39.70	20.82	22.02	24.73	28.37	16.37
RigL	39.96	41.33	45.34	51.49	25.85	66.35	37.18	49.39	149.17
CHT	31.02	32.99	35.01	41.87	21.02	22.82	26.27	30.01	–
CHTs	**28.12**	**29.84**	**33.03**	**36.47**	**20.10**	**21.33**	**23.71**	**26.45**	**14.53***
MEST	28.26	29.94	33.60	37.87	21.32	22.21	24.98	27.96	60.36
GMP	29.22	30.59	33.68	39.00	20.49	22.28	23.61	27.16	31.76
GraNet	30.55	31.51	33.76	39.98	22.84	29.03	26.81	61.31	79.44
CHTss	**27.62**	**29.00**	**31.42**	**35.10**	**19.85**	**20.70**	**22.51**	**25.07**	**15.41**

This work was supported by: The Zhou Yahui Chair Professorship award of Tsinghua University (to CVC); The National High-Level Talent Program of the Ministry of Science and Technology of China (grant number 20241710001, to CVC).
The Center for Information Services and High Performance Computing (ZIH) of the TU Dresden and ScaDS.AI for providing HPC resources.